

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANTS:

- (A) Cooke, David
- (B) Debet, Martine
- (C) Gidley, Micheal John
- (D) Jobling, Stephen Alan
- (E) Safford, Richard
- (F) Sidebottom, Christopher Michael
- (G) Westcott, Roger John

(ii) TITLE OF INVENTION: Improvements in or Relating to Plant Starch Composition

(iii) NUMBER OF SEQUENCES: 20

## (iv) CORRESPONDANCE ADDRESS:

- (A) NAME: National Starch and Chemical Company
- (B) STREET: 10 Finderne Avenue, P.O. Box 6500
- (C) CITY: Bridgewater
- (D) STATE: New Jersey
- (E) COUNTRY: United States
- (F) ZIP CODE: 08807-500

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: IMB 1.44 MB High Density Diskette
- (B) COMPUTER: COMPAQ Deskpro 590 (IBM PC compatible)
- (C) OPERATING SYSTEM: WINDOWS 95
- (D) SOFTWARE: Word 7.0 for Windows

## (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE: Filed concurrently herewith
- (C) CLASSIFICATION

## (vii) PRIOR APPLICATION DATA

- (A) APPLICATION NUMBER: PCT/GB96/01075
- (B) INTERNATIONAL FILING DATE: May 3, 1996
- (C) PRIORITY DATE: May 5, 1995

## (viii) ATTORNEY INFORMATION

- (A) NAME: Karen G. Kaiser
- (B) REGISTRATION NO: 33,506
- (C) DOCKET NUMBER: 1627

## (ix) TELECOMMUNICATION INFORMATION

- (A) TELEPHONE: (908) 575-6152
- (B) FACSIMILE: (908) 707-3706
- (C) E-MAIL: KAREN.KAISER@NSTARCH.COM

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AAGGATCCGT CGACATCGAT AATACGACTC ACTATAGGGA TTTT TTTT TTTT

57

## (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

AAGGATCCGT CGACATC

17

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GACATCGATA ATACGAC

17

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CATCCAACCA CCATCTCGCA

20

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TTGAGAGAAG ATACCTAAGT

20

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATGTTCACTC CATCTAAAGT

20

## (2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

AGAACAACAA TTCCTAGCTC

20

## (2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGGCCTGA ACTCAGCAAT

20

## (2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGTCCCAGCA TTCGACATAA

20

## (2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 26 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTTGGATCCT TGAACTCAGC AATTGT

26

## (2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 29 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TAACTCGAGC AACGCGATCA CAAGTTCGT

29

## (2) INFORMATION FOR SEQ ID NO: 12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3003 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GATGGGGCCT TGAACCTCAGC AATTTGACAC TCAGTTAGTT ACACTGCCAT CACTTATCAG	60
ATCTCTATTT TTTCTCTTAA TTCCAACCAA GGAATGAATA AAAAGATAGA TTTGTAAAAA	120
CCCTAAGGAG AGAAGAAGAA AGATGGTGTA TACACTCTCT GGAGTTTCGTT TTCCTACTGT	180
TCCATCAGTG TACAAATCTA ATGGATTGAG CAGTAATGGT GATCGGAGGA ATGCTAATAT	240
TTCTGTATTC TTGAAAAAAC ACTCTCTTTC ACGGAAGATC TTGGCTGAAA AGTCTTCTTA	300
CAATTCCGAA TCCCGACCTT CTACAATTGC AGCATCGGGG AAAGTCCTTG TGCCTGGAAT	360
CCAGAGTGAT AGCTCCTCAT CCTCAACAGA TCAATTGAG TTGCTGAGA CATCTCCAGA	420
AAATTCCCCA GCATCAACTG ATGTAGATAG TTCAACAATG GAACACGCTA GCCAGATTAA	480
AACTGAGAAC GATGACGTTG AGCCGTCAAG TGATCTTACA GGAAGTGTG AAGAGCTGGA	540
TTTTGCTTCA TCACTACAAC TACAAGAAGG TGGTAACTG GAGGAGTCTA AAACATTAAA	600
TACTTCTGAA GAGACAATTA TTGATGAATC TGATAGGATC AGAGAGAGGG GCATCCCTCC	660
ACCTGGACTT GGTGAGAAGA TTTATGAAAT AGACCCCTT TTGACAACT ATCGTCAACA	720
CCTTGATTAC AGGTATTCAC AGTACAAGAA ACTGAGGGAG GCAATTGACA AGTATGAGGG	780
TGGTTTGGA GCTTTTTCTC GTGGTTATGA AAGAATGGGT TCACTCGTA GTGCTACAGG	840
TATCACTTAC CGTGAGTGGG CTCTGCTGTC CCAGTCAGCT GCCCTCATG GGGATTTCAA	900
CAATTGGGAC GCAAATGCTG ACTTTATGAC TCGGAATGAA TTTGGTGTCT GAGAGATTTT	960
TCTGCCAAAT AATGTGGATG GTTCTCCTGC AATTCCTCAT GGGTCCAGAG TGAAGATACG	1020
TATGGACACT CCATCAGGTG TTAAGGATTC CATTCCTGCT TGGATCAACT ACTCTTTACA	1080
GCTTCCTGAT GAAATTCAT ATAATGGAAT ATATTATGAT CCACCCGAAG AGGAGAGGTA	1140
TATCTTCCAA CACCCACGGC CAAAGAAACC AAAGTCGGTG AGAATATATG AATCTCATAT	1200
TGGAATGAGT AGTCCGGAGC CTAATAATTAA CTCATACGTG AATTTTAGAG ATGAAGTTCT	1260
TCCTCGCATA AAAAAAGCTT GGGTACAATG CGGTGCAAAT TATGGCTATT CAAGAGCATT	1320
CTTATTATGC TAGTTTTGGT TATCATGTCA CAAATTTTTT TGCACCAAGC AGCCGTTTTG	1380
GAACGCCCCG CGACCTTAAG TCTTTGATTG ATAAAGCTCA TGAGCTAGGA ATTGTTGTTT	1440
TCATGGACAT TGTTACAGC CATGCATCA ATAATACTTT AGATGGACTG AACATGTTTG	1500
ACGGCACAGA TAGTTGTTAC TTTCACTCTG GAGCTCGTGG TTATCATTGG ATGTGGGATT	1560

TCCGCCTCTT	TAACATATGGA	AACTGGGAGG	TACTTAGGTA	TCTTCTCTCA	AATGCGAGAT	1620
GGTGGTTGGA	TGAGTTCAAA	TTTGATGGAT	TTAGATTGTA	TGGTGTGACA	TCAATGATGT	1680
GTA CTACCA	CGGATTATCG	GTGGGATTCA	CTGGGAACTA	CGAGGAATAC	TTTGGACTCG	1740
CAACTGATGT	GGATGCTGTT	GTGTATCTGA	TGCTGGTCAA	CGATCTTATT	CATGGGCTTT	1800
TCCCAGATGC	AATTACCATT	GGTGAAGATG	TTAGCGGAAT	GCCGACATTT	TGTGTTCCCCG	1860
TTCAAGATGG	GGGTGTTGGC	TTTGACTATC	GGCTGCATAT	GGCAATTGCT	GATAAATGGA	1920
TTGAGTTGCT	CAAGAAACGG	GATGAGGATT	GGAGAGTGGG	TGATATTGTT	CATACACTGA	1980
CAAATAGAAG	ATGGTCGGAA	AAGTGTGTTT	CATACGCTGA	AAGTCATGAT	CAAGCTCTAG	2040
TCGGTGATAA	AACTATAGCA	TTCTGGCTGA	TGGACAAGGA	TATGTATGAT	TTTATGGCTC	2100
TGGATAGACC	GTCAACATCA	TTAATAGATC	GTGGGATAGC	ATTACACAAG	ATGATTAGGC	2160
TTGTAACAT	GGGATTAGGA	GGAGAAGGGT	ACCTAAATTT	CATGGGAAAT	GAATTCGGCC	2220
ACCCTGAGTG	GATTGATTTC	CCTAGGGCTG	AACAACACCT	CTCTGATGGC	TCAGTAATTC	2280
CCAGAAACCA	ATTCAGTTAT	GATAAATGCA	GACGGAGATT	TGACCTGGGA	GATGCAGAAT	2340
ATTTAAGATA	CCGTGGGTG	CAAGAATTTG	ACCGGGCTAT	GCAGTATCTT	GAAGATAAAT	2400
ATGAGTTTAT	GACTTCAGAA	CACCAGTTCA	TATCACGAAA	GGATGAAGGA	GATAGGATGA	2460
TTGTATTGTA	AAAAGGAAAC	CTAGTTTTTG	TCTTTAATTT	TCACTGGACA	AAAGGCTATT	2520
CAGACTATCG	CATAGGCTGC	CTGAAGCCTG	GAAATACAA	GGTTGCCTTG	GACTCAGATG	2580
ATCCACTTTT	TGGTGGCTTC	GGGAGAATTG	ATCATAATGC	CGAATATTTC	ACCTTTGAAG	2640
GATGGTATGA	TGATCGTCCT	CGTTCAATTA	TGGTGTATGC	ACCTAGTAGA	ACAGCAGTGG	2700
TCTATGCACT	AGTAGACAAA	GAAGAAGAAG	AAGAAGAAGA	AGTAGCAGTA	GTAGAAGAAG	2760
TAGTAGTAGA	AGAAGAATGA	ACGAACCTGT	GATCGCGTTG	AAAGATTGTA	ACGCCACATA	2820
GAGCTTCTTG	ACGTATCTGG	CAATATTGCA	TTAGTCTTGG	CGGAATTTC	TGTGACAACA	2880
GGTTTGCAAT	TCTTTCCACT	ATTAGTAGTG	CAACGATATA	CGCAGAGATG	AAGTGCTGAA	2940
CAAAAACATA	TGTA AAATCG	ATGAATTTAT	GTGCAATGCT	GGGACGATCG	AATTCCTGCA	3000
GCC						3003

## (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2975 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TTGATGGGCC	TTGAACTCAG	CAATTTGACA	CTCAGTTAGT	TACACTCCTA	TCACTTATCA	60
GATCTCTATT	TTTCTCTT	ATTCCAACCA	GGGGAATGAA	TAAAAGGATA	GATTTGTAAA	120

AACCCTAAGG AGAGAAGAAG AAAGATGGTG TATATACTCT CTGGAGTTCG TTTTCCTACT	180
GTTCATCAG TGTACAAATC TAATGGATTG AGCAGTAATG GTGATCGGAG GAATGCTAAT	240
GTTCCTGTAT TCTTGAAAAA GCACCTCTCTT TCACGGAAGA TCTTGGCTGA AAAGTCTTCT	300
TACAATTCCG AATTCGGACC TTCTACAGTT GCAGCATCGG GGAAAGTCCT TGTGCCTGGA	360
ACCCAGAGTG ATAGCTCCTC ATCCCTCAACA GACCAATTTG AGTTCACTGA GACATCTCCA	420
GAAAATTCCC CAGCATCAAC TGATGTAGAT AGTTCAACAA TGGAACACGC TAGCCAGATT	480
AAAAGTGAAG ACGATGACGT TGAGCCGTCA AGTGATCTTA CAGGAAGTGT TGAAGAGCTG	540
GATTTTGCTT CATCACTACA ACTACAAGAA GGTGGTAAAC TGGAGGAGTC TAAACATTA	600
AATACTTCTG AAGAGACAAT TATTGATGAA TCTGATAGGA TCAGAGAGAG GGCATCCCT	660
CCACCTGGAC TTGGTCAGAA GATTATGAA ATAGACCCCC TTTTGACAAA CTATCGTCAA	720
CACCTTGATT ACAGGTATTC ACAGTACAAG AAAGTGGGG AGGCAATTGA CAAGTATGAG	780
GGTGGTTTGG AAGCTTTTCT CGTGGTTATG AAAAAATGGG TTTCACCTCGT AGTGCTACAG	840
GTATCACTTA CCGTGAGTGG GCTCCTGGTG CCCAGTCAGC TGCCCTCATT GGAGATTTCA	900
ACAATTGGGA CGCAATGCT GACATTATGA CTCGGAATGA ATTTGGTGTC TGGGAGATTT	960
TTCTGCCAAA TAATGTGGAT GGTTCCTCTG CAATTCCTCA TGGGTCCAGA GTGAAGATAC	1020
GTATGGACAC TCCATCAGGT GTTAAGGATT CCATTCCTGC TTGGATCAAC TACTCTTTAC	1080
AGCTTCCTGA TGAAATTCCA TATAATGGAA TATATTATGA TCCACCCGAA GAGGAGAGGT	1140
ATATCTTCCA ACACCCACGG CCAAGAAAC CAAAGTCGCT GAGAATATAT GAATCTCATA	1200
TTGGAATGAG TAGTCCGGAG CCTAAAATTA ACTCATACGT GAATTTTAGA GATGAAGTTC	1260
TTCTCGCAT AAAAAAGCTT GGGTACAATG CGCTGCGAAT TATGGCTATT CAAGAGCATT	1320
CTTATTATGC TAGTTTTGGT TATCATGTCA CAAATTTTTT TGCACCAAGC AGCCGTTTTG	1380
GAACGCCCGA CGACCTTAAG TCTTCGATG ATAAAGCTCA TGAGCTAGGA ATTGTTGTTC	1440
TCATGGACAT CGTTCACAGC CATGCATCAA ATAATACTTT AGATGGACTG AACATGTTTG	1500
ACGGCACCGA TAGTTGTTAC TTTCACCTCTG GAGCTCGTGG TTATCATTGG ATGTGGGATT	1560
CCGCCTCTTT AACTATGGAA ACTGGGAGGT ACTTAGGTAT CTCTCTCTAA ATGCGAGATG	1620
GTGGTTGGAT GAGTTCAAAT TTGATGGATT TAGATTCGAT GGTGTGACAT CAATGATGTA	1680
TACTCACCAC GGATTATCGG TGGGATTAC TGGGAACAC GAGGAATACT TTGGACTCGC	1740
AACTGATGTG GATGCTGTTG TGTATCTGAT GCTGGTCAAC GATCTTATTC ATAGGCTTTT	1800
CCCAGATGCA ATTACCATTG GTGAAGATGT TAGCGGAATG CCGACATTTT GTATTCCTGT	1860
TCAAGATGGG GGTGTTGGCT TTGACTATCG GCTGCATATG GCAATTGCTG ATAAATGGAT	1920
TGAGTTGCTC AAGAAACGGG ATGAGGATG GAGAGTGGGT GATATTGTTC ATACACTGAC	1980

AAATAGAAGA TGGTCGGAAG AGTGTGTTTC ATACGCTGAA AGTCATGATC AAGCTCTAGT	2040
CGGTGATAAA ACTATAGCAT TCTGGCTGAT GGACAAGGAT ATGTATGATT TTATGGCTCT	2100
GGATAGACCG CCAACATCAT TAATAGATCG TGGGATAGCA TTGCACAAGA TGATTAGGCT	2160
TGTAACATATG GGATTAGGAG GAGAAGGGTA CCTAAATTTT ATGGGAAAATG AATTCGGCCA	2220
CCCTGAGTGG ATTGATTTCC CTAGGGCTGA GCCACACCTT TCTGATGGCT CAGTAATTC	2280
CGGAAACCAA TTCAGTTATG ATAAATGCAG ACGGAGATTT GACCTGGGAG ATGCAGAATA	2340
TTTAAGATAC CATGGGTAC AAGAATTTGA CTGGGCTATG CAGTATCTTG AAGATAAATA	2400
TGAGTTTATG ACTTCAGAAC ACCAGTTCAT ATCACGAAAG GATGAAGGAG ATAGGATGAT	2460
TGTATTTGAA AGAGGAAACC TAGTTTTCGT CTTTAATTTT CACTGGACAA ATAGCTATTC	2520
AGACTATCGC ATAGCTGCC TGAAGCCTGG AAAATACAAG GTTGTCTTGG ACTCAGATGA	2580
TCCACTTTTT GGTGGCTTCG GGAGAATTGA TCATAATGCC GAATATTTCA CCTCTGAAGG	2640
ATCGTATGAT GATCGTCCTT GTTCAATTAT GGTGTATGCA CCTAGTAGAA CAGCAGTGGT	2700
CTATGCACTA GTAGACAAAC TAGAAGTAGC AGTAGTAGAA GAACCCATTG AAGAATGAAC	2760
GAAC TTGTGA TCGCGTTGAA AGATTGAAC GTTACTTGGT CATCCACATA GAGCTTCTTG	2820
ACATCAGTCT TGGCGGAATT GCATGTGACA ACAAGGTTTG CAGTTCTTTC CACTATTAGT	2880
AGTCCACCGA TATACGCAGA GATGAAGTGC TGAACAAACA TATGTAAAAT CGATGAATTT	2940
ATGTCGAATG CTGGGACGAT CGAATTCCTG CAGCC	2975

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3033 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 145..2790

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TTGATGGGGC CTTGAATCA GCAATTGAC ACTCAGTTAG TTACACTCCT ATCACTTATC	60
AGATCTCTAT TTTTCTCTT AATTCCAACC AAGGAATGAA TAAAGGATA GATTGTAA	120
AACCCTAAGG AGAGAAGAAG AAAG ATG GTG TAT ACA CTC TCT GGA GTT CGT	171
Met Val Tyr Thr Leu Ser Gly Val Arg	
1 5	
TTT CCT ACT GTT CCA TCA GTG TAC AAA TCT AAT GGA TTC AGC AGT AAT	219
Phe Pro Thr Val Pro Ser Val Tyr Lys Ser Asn Gly Phe Ser Ser Asn	
10 15 20 25	
GGT GAT CGG AGG AAT GCT AAT GTT TCT GTA TTC TTG AAA AAG CAC TCT	267

Gly	Asp	Arg	Arg	Asn	Ala	Asn	Val	Ser	Val	Phe	Leu	Lys	Lys	His	Ser	
				30					35					40		
CTT	TCA	CGG	AAG	ATC	TTG	GCT	GAA	AAG	TCT	TCT	TAC	AAT	TCC	GAA	TTC	315
Leu	Ser	Arg	Lys	Ile	Leu	Ala	Glu	Lys	Ser	Ser	Tyr	Asn	Ser	Glu	Phe	
			45					50					55			
CGA	CCT	TCT	ACA	GTT	GCA	GCA	TCG	GGG	AAA	GTC	CTT	GTG	CCT	GGA	ACC	363
Arg	Pro	Ser	Thr	Val	Ala	Ala	Ser	Gly	Lys	Val	Leu	Val	Pro	Gly	Thr	
			60					65				70				
CAG	AGT	GAT	AGC	TCC	TCA	TCC	TCA	ACA	GAC	CAA	TTT	GAG	TTC	ACT	GAG	411
Gln	Ser	Asp	Ser	Ser	Ser	Ser	Ser	Thr	Asp	Gln	Phe	Glu	Phe	Thr	Glu	
			75				80					85				
ACA	TCT	CCA	GAA	AAT	TCC	CCA	GCA	TCA	ACT	GAT	GTA	GAT	AGT	TCA	ACA	459
Thr	Ser	Pro	Glu	Asn	Ser	Pro	Ala	Ser	Thr	Asp	Val	Asp	Ser	Ser	Thr	
			90			95				100					105	
ATG	GAA	CAC	GCT	AGC	CAG	ATT	AAA	ACT	GAG	AAC	GAT	GAC	GTT	GAG	CCG	507
Met	Glu	His	Ala	Ser	Gln	Ile	Lys	Thr	Glu	Asn	Asp	Asp	Val	Glu	Pro	
				110					115					120		
TCA	AGT	GAT	CTT	ACA	GGA	AGT	GTT	GAA	GAG	CTG	GAT	TTT	GCT	TCA	TCA	555
Ser	Ser	Asp	Leu	Thr	Gly	Ser	Val	Glu	Glu	Leu	Asp	Phe	Ala	Ser	Ser	
			125					130					135			
CTA	CAA	CTA	CAA	GAA	GGT	GGT	AAA	CTG	GAG	GAG	TCT	AAA	ACA	TTA	AAT	603
Leu	Gln	Leu	Gln	Glu	Gly	Gly	Lys	Leu	Glu	Glu	Ser	Lys	Thr	Leu	Asn	
			140				145					150				
ACT	TCT	GAA	GAG	ACA	ATT	ATT	GAT	GAA	TCT	GAT	AGG	ATC	AGA	GAG	AGG	651
Thr	Ser	Glu	Glu	Thr	Ile	Ile	Asp	Glu	Ser	Asp	Arg	Ile	Arg	Glu	Arg	
			155			160					165					
GGC	ATC	CCT	CCA	CCT	GGA	CTT	GGT	CAG	AAG	ATT	TAT	GAA	ATA	GAC	CCC	699
Gly	Ile	Pro	Pro	Pro	Gly	Leu	Gly	Gln	Lys	Ile	Tyr	Glu	Ile	Asp	Pro	
					175					180					185	
CTT	TTG	ACA	AAC	TAT	CGT	CAA	CAC	CTT	GAT	TAC	AGG	TAT	TCA	CAG	TAC	747
Leu	Leu	Thr	Asn	Tyr	Arg	Gln	His	Leu	Asp	Tyr	Arg	Tyr	Ser	Gln	Tyr	
				190					195					200		
AAG	AAA	CTG	AGG	GAG	GCA	ATT	GAC	AAG	TAT	GAG	GGT	GGT	TTG	GAA	GCC	795
Lys	Lys	Leu	Arg	Glu	Ala	Ile	Asp	Lys	Tyr	Glu	Gly	Gly	Leu	Glu	Ala	
			205					210					215			
TTT	TCT	CGT	GGT	TAT	GAA	AAA	ATG	GGT	TTC	ACT	CGT	AGT	GCT	ACA	GGT	843
Phe	Ser	Arg	Gly	Tyr	Glu	Lys	Met	Gly	Phe	Thr	Arg	Ser	Ala	Thr	Gly	
			220				225					230				
ATC	ACT	TAC	CGT	GAG	TGG	GCT	CTT	GGT	GCC	CAG	TCA	GCT	GCC	CTC	ATT	891
Ile	Thr	Tyr	Arg	Glu	Trp	Ala	Leu	Gly	Ala	Gln	Ser	Ala	Ala	Leu	Ile	
			235			240					245					
GGA	GAT	TTC	AAC	AAT	TGG	GAC	GCA	AAT	GCT	GAC	ATT	ATG	ACT	CGG	AAT	939
Gly	Asp	Phe	Asn	Asn	Trp	Asp	Ala	Asn	Ala	Asp	Ile	Met	Thr	Arg	Asn	
					255				260						265	
GAA	TTT	GGT	GTC	TGG	GAG	ATT	TTT	CTG	CCA	AAT	AAT	GTG	GAT	GGT	TCT	987
Glu	Phe	Gly	Val	Trp	Glu	Ile	Phe	Leu	Pro	Asn	Asn	Val	Asp	Gly	Ser	



270										275										280										
CCT Pro	GCA Ala	ATT Ile	CCT Pro 285	CAT His	GGG Gly	TCC Ser	AGA Arg	GTG Val 290	AAG Lys	ATA Ile	CGT Arg	ATG Met 295	GAC Asp	ACT Thr	CCA Pro	1035														
TCA Ser	GGT Gly 300	GTT Val	AAG Lys	GAT Asp	TCC Ser	ATT Ile	CCT Pro 305	GCT Ala	TGG Trp	ATC Ile	AAC Asn	TAC Tyr 310	TCT Ser	TTA Leu	CAG Gln	1083														
CTT Leu	CCT Pro 315	GAT Asp	GAA Glu	ATT Ile	CCA Pro	TAT Tyr 320	AAT Asn	GGA Gly	ATA Ile	CAT His	TAT Tyr 325	GAT Asp	CCA Pro	CCC Pro	GAA Glu	113														
GAG Glu 330	GAG Glu	AGG Arg	TAT Tyr	ATC Ile	TTC Phe 335	CAA Gln	CAC His	CCA Pro	CGG Arg	CCA Pro 340	AAG Lys	AAA Lys	CCA Pro	AAG Lys	TCG Ser 345	1179														
CTG Leu	AGA Arg	ATA Ile	TAT Tyr	GAA Glu 350	TCT Ser	CAT His	ATT Ile	GGA Gly 355	ATG Met	AGT Ser	AGT Ser	CCG Pro	GAG Glu	CCT Pro 360	AAA Lys	1227														
ATT Ile	AAC Asn	TCA Ser	TAC Tyr 365	GTG Val	AAT Asn	TTT Phe	AGA Arg	GAT Asp 370	GAA Glu	GTT Val	CTT Leu	CCT Pro 375	CGC Arg	ATA Ile	AAA Lys	1275														
AAG Lys	CTT Leu	GGG Gly 380	TAC Tyr	AAT Asn	GCG Ala	CTG Leu	CAA Gln 385	ATT Ile	ATG Met	GCT Ala	ATT Ile	CAA Gln 390	GAG Glu	CAT His	TCT Ser	1323														
TAT Tyr	TAC Tyr 395	GCT Ala	AGT Ser	TTT Phe	GGT Gly 400	TAT Tyr	CAT His	GTC Val	ACA Thr	AAT Asn	TTT Phe 405	TTT Phe	GCA Ala	CCA Pro	AGC Ser	1371														
AGC Ser 410	CGT Arg	TTT Phe	GGA Gly	ACG Thr	CCC Pro 415	GAC Asp	GAC Asp	CTT Leu	AAG Lys	TCT Ser 420	TTG Leu	ATT Ile	GAT Asp	AAA Lys	GCT Ala 425	1419														
CAT His	GAG Glu	CTA Leu	GGA Gly	ATT Ile 430	GTT Val	GTT Val	CTC Leu	ATG Met	GAC Asp 435	ATT Ile	GTT Val	CAC His	AGC Ser	CAT His 440	GCA Ala	1467														
TCA Ser	AAT Asn	AAT Asn	ACT Thr 445	TTA Leu	GAT Asp	GGA Gly	CTG Leu	AAC Met 450	ATG Met	TTT Phe	GAC Asp	TGC Cys	ACC Thr 455	GAT Asp	AGT Ser	1515														
TGT Cys	TAC Tyr	TTT Phe 460	CAC His	TCT Ser	GGA Gly	GCT Ala	CGT Arg 465	GGT Gly	TAT Tyr	CAT His	TGG Trp	ATG Met 470	TGG Trp	GAT Asp	TCC Ser	1563														
CGC Arg	CTC Leu 475	TTT Phe	AAC Asn	TAT Tyr	GGA Gly	AAC Asn 480	TGG Trp	GAG Glu	GTA Val	CTT Leu	AGG Arg 485	TAT Tyr	CTT Leu	CTC Leu	TCA Ser	1611														
AAT Asn 490	GCG Ala	AGA Arg	TGG Trp	TGG Trp	TTG Leu 495	GAT Asp	GCG Ala	TTC Phe	AAA Lys	TTT Phe 500	GAT Asp	GGA Gly	TTT Phe	AGA Arg	TTT Phe 505	1659														
GAT Asp	GGT Gly	GTG Val	ACA Thr	TCA Ser 510	ATG Met	ATG Met	TAT Tyr	ATT Ile	CAC His 515	CAC His	GGA Gly	TTA Leu	TCG Ser	GTG Val 520	GGA Gly	1707														
TTC	ACT	GGG	AAC	TAC	GAG	GAA	TAC	TTT	GGA	CTC	GCA	ACT	GAT	GTG	GAT	1755														

Phe	Thr	Gly	Asn	Tyr	Glu	Glu	Tyr	Phe	Gly	Leu	Ala	Thr	Asp	Val	Asp	
			525					530					535			
GCT	GTT	GTG	TAT	CTG	ATG	CTG	GTC	AAC	GAT	CTT	ATT	CAT	GGG	CTT	TTC	1803
Ala	Val	Val	Tyr	Leu	Met	Leu	Val	Asn	Asp	Leu	Ile	His	Gly	Leu	Phe	
		540					545					550				
CCA	GAT	GCA	ATT	ACC	ATT	GGT	GAA	GAT	GTT	AGC	GGA	ATG	CCG	ACA	TTT	1851
Pro	Asp	Ala	Ile	Thr	Ile	Gly	Glu	Asp	Val	Ser	Gly	Met	Pro	Thr	Phe	
	555					560					565					
TGT	ATT	CCC	GTC	CAA	GAG	GGG	GGT	GTT	GGC	TTT	GAC	TAT	CGG	CTG	CAT	1899
Cys	Ile	Pro	Val	Gln	Glu	Gly	Gly	Val	Gly	Phe	Asp	Tyr	Arg	Leu	His	
570					575					580					585	
ATG	GCA	ATT	GCT	GAT	AAA	CGG	ATT	GAG	TTG	CTC	AAG	AAA	CGG	GAT	GAG	1947
Met	Ala	Ile	Ala	Asp	Lys	Arg	Ile	Glu	Leu	Lys	Lys	Arg	Arg	Asp	Glu	
				590					595					600		
GAT	TGG	AGA	GTG	GGT	GAT	ATT	GTT	CAT	ACA	CTG	ACA	AAT	AGA	AGA	TGG	1995
Asp	Trp	Arg	Val	Gly	Asp	Ile	Val	His	Thr	Leu	Thr	Asn	Arg	Arg	Trp	
			605					610					615			
TCG	GAA	AAG	TGT	GTT	TCA	TAC	GCT	GAA	AGT	CAT	GAT	CAA	GCT	CTA	GTC	2043
Ser	Glu	Lys	Cys	Val	Ser	Tyr	Ala	Glu	Ser	His	Asp	Gln	Ala	Leu	Val	
		620					625					630				
GGT	GAT	AAA	ACT	ATA	GCA	TTC	TGG	CTG	ATG	GAC	AAG	GAT	ATG	TAT	GAT	2091
Gly	Asp	Lys	Thr	Ile	Ala	Phe	Trp	Leu	Met	Asp	Lys	Asp	Met	Tyr	Asp	
	635					640					645					
TTT	ATG	GCT	CTG	GAT	AGA	CCG	TCA	ACA	TCA	TTA	ATA	GAT	CGT	GGG	ATA	2139
Phe	Met	Ala	Leu	Asp	Arg	Pro	Ser	Thr	Ser	Leu	Ile	Asp	Arg	Gly	Ile	
650					655					660					665	
GCA	TTG	CAC	AAG	ATG	ATT	AGG	CTT	GTA	ACT	ATG	GGA	TTA	GGA	GGA	GAA	2187
Ala	Leu	His	Lys	Met	Ile	Arg	Leu	Val	Thr	Met	Gly	Leu	Gly	Gly	Glu	
				670					675					680		
GGG	TAC	CTA	AAT	TTC	ATG	GGA	AAT	GAA	TTC	GGC	CAC	CCT	GAG	TGG	ATT	2235
Gly	Tyr	Leu	Asn	Phe	Met	Gly	Asn	Glu	Phe	Gly	His	Pro	Glu	Trp	Ile	
			685					690					695			
GAT	TTC	CCT	AGG	GCT	GAA	CAA	CAC	CTC	TCT	GAT	GGC	TCA	GTA	ATC	CCC	2283
Asp	Phe	Pro	Arg	Ala	Glu	Gln	His	Leu	Ser	Asp	Gly	Ser	Val	Ile	Pro	
		700					705					710				
GGA	AAC	CAA	TTC	AGT	TAT	GAT	AAA	TGC	AGA	CGG	AGA	TTT	GAC	CTG	GGA	2331
Gly	Asn	Gln	Phe	Ser	Tyr	Asp	Lys	Cys	Arg	Arg	Arg	Phe	Asp	Leu	Gly	
	715					720					725					
GAT	GCA	GAA	TAT	TTA	AGA	TAC	CGT	GGG	TTG	CAA	GAA	TTT	GAC	CGG	CCT	2379
Asp	Ala	Glu	Tyr	Leu	Arg	Tyr	Arg	Gly	Leu	Gln	Glu	Phe	Asp	Arg	Pro	
730					735				740						745	
ATG	CAG	TAT	CTT	GAA	GAT	AAA	TAT	GAG	TTT	ATG	ACT	TCA	GAA	CAC	CAG	2427
Met	Gln	Tyr	Leu	Glu	Asp	Lys	Tyr	Glu	Phe	Met	Thr	Ser	Glu	His	Gln	
				750				755						760		
TTC	ATA	TCA	CGA	AAG	GAT	GAA	GGA	GAT	AGG	ATG	ATT	GTA	TTT	GAA	AAA	2475
Phe	Ile	Ser	Arg	Lys	Asp	Glu	Gly	Asp	Arg	Met	Ile	Val	Phe	Glu	Lys	
			765					770					775			

GGA AAC CTA GTT TTT GTC TTT AAT TTT CAC TGG ACA AAA AGC TAT TCA Gly Asn Leu Val Phe Val Phe Asn Phe His Trp Thr Lys Ser Tyr Ser 780 785 790	2523
GAC TAT CGC ATA GCC TGC CTG AAG CCT GGA AAA TAC AAG GTT GCC TTG Asp Tyr Arg Ile Ala Cys Leu Lys Pro Gly Lys Tyr Lys Val Ala Leu 795 800 805	2571
GAC TCA GAT GAT CCA CTT TTT GGT GGC TTC GGG AGA ATT GAT CAT AAT Asp Ser Asp Asp Pro Leu Phe Gly Gly Phe Gly Arg Ile Asp His Asn 810 815 820 825	2619
GCC GAA TAT TTC ACC TTT GAA GGA TGG TAT GAT GAT CGT CCT CGT TCA Ala Glu Tyr Phe Thr Phe Glu Gly Trp Tyr Asp Asp Arg Pro Arg Ser 830 835 840	2667
ATT ATG GTG TAT GCA CCT TGT AAA ACA GCA GTG GTC TAT GCA CTA GTA Ile Met Val Tyr Ala Pro Cys Lys Thr Ala Val Val Tyr Ala Leu Val 845 850 855	2715
GAC AAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GTA GCA GCA Asp Lys Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Val Ala Ala 860 865 870	2763
GTA GAA GAA GTA GTA GTA GAA GAA GAA TGAACGAACT TGTGATCGCG Val Glu Glu Val Val Val Glu Glu Glu 875 880	2810
TTGAAAGATT TGAACGCTAC ATAGAGCTTC TTGACGTATC TGGCAATATT GCATCAGTCT	2870
TGGCGGAATT TCATGTGACA CAAGGTTTGC AATTCTTTCC ACTATTAGTA GTGCAACGAT	2930
ATACGCAGAG ATGAAGTGCT GAACAAACAT ATGTAAAATC GATGAATTTA TGTGCAATGC	2990
TGGGACGATC GAATTCCTGC AGGCCGGGGG ACCCCTTAGT TCT	3033

## (2) INFORMATION FOR SEQ ID NO: 15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 882 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Met Val Tyr Thr Leu Ser Gly Val Arg Phe Pro Thr Val Pro Ser Val  
1 5 10 15

Tyr Lys Ser Asn Gly Phe Ser Ser Asn Gly Asp Arg Arg Asn Ala Asn  
20 25 30

Val Ser Val Phe Leu Lys Lys His Ser Leu Ser Arg Lys Ile Leu Ala  
35 40 45

Glu Lys Ser Ser Tyr Asn Ser Glu Phe Arg Pro Ser Thr Val Ala Ala  
50 55 60

Ser Gly Lys Val Leu Val Pro Gly Thr Gln Ser Asp Ser Ser Ser Ser  
65 70 75 80

Ser Thr Asp Gln Phe Glu Phe Thr Glu Thr Ser Pro Glu Asn Ser Pro

Ala	Ser	Thr	Asp	Val	Asp	Ser	Ser	Thr	Met	Glu	His	Ala	Ser	Gln	Ile
			100					105					110		
Lys	Thr	Glu	Asn	Asp	Asp	Val	Glu	Pro	Ser	Ser	Asp	Leu	Thr	Gly	Ser
		115					120					125			
Val	Glu	Glu	Leu	Asp	Phe	Ala	Ser	Ser	Leu	Gln	Leu	Gln	Glu	Gly	Gly
		130				135					140				
Lys	Leu	Glu	Glu	Ser	Lys	Thr	Leu	Asn	Thr	Ser	Glu	Glu	Thr	Ile	Ile
	145				150					155				160	
Asp	Glu	Ser	Asp	Arg	Ile	Arg	Glu	Arg	Gly	Ile	Pro	Pro	Pro	Gly	Leu
				165					170					175	
Gly	Gln	Lys	Ile	Tyr	Glu	Ile	Asp	Pro	Leu	Leu	Thr	Asn	Tyr	Arg	Gln
			180					185					190		
His	Leu	Asp	Tyr	Arg	Tyr	Ser	Gln	Tyr	Lys	Lys	Leu	Arg	Glu	Ala	Ile
		195					200					205			
Asp	Lys	Tyr	Glu	Gly	Gly	Leu	Glu	Ala	Phe	Ser	Arg	Gly	Tyr	Glu	Lys
	210					215					220				
Met	Gly	Phe	Thr	Arg	Ser	Ala	Thr	Gly	Ile	Thr	Tyr	Arg	Glu	Trp	Ala
	225				230					235				240	
Leu	Gly	Ala	Gln	Ser	Ala	Ala	Leu	Ile	Gly	Asp	Phe	Asn	Asn	Trp	Asp
				245					250					255	
Ala	Asn	Ala	Asp	Ile	Met	Thr	Arg	Asn	Glu	Phe	Gly	Val	Trp	Glu	Ile
			260					265					270		
Phe	Leu	Pro	Asn	Asn	Val	Asp	Gly	Ser	Pro	Ala	Ile	Pro	His	Gly	Ser
		275					280					285			
Arg	Val	Lys	Ile	Arg	Met	Asp	Thr	Pro	Ser	Gly	Val	Lys	Asp	Ser	Ile
	290					295					300				
Pro	Ala	Trp	Ile	Asn	Tyr	Ser	Leu	Gln	Leu	Pro	Asp	Glu	Ile	Pro	Tyr
	305				310					315				320	
Asn	Gly	Ile	His	Tyr	Asp	Pro	Pro	Glu	Glu	Arg	Tyr	Ile	Phe	Gln	
				325				330					335		
His	Pro	Arg	Pro	Lys	Lys	Pro	Lys	Ser	Leu	Arg	Ile	Tyr	Glu	Ser	His
			340					345					350		
Ile	Gly	Met	Ser	Ser	Pro	Glu	Pro	Lys	Ile	Asn	Ser	Tyr	Val	Asn	Phe
		355					360					365			
Arg	Asp	Glu	Val	Leu	Pro	Arg	Ile	Lys	Lys	Leu	Gly	Tyr	Asn	Ala	Leu
	370					375					380				
Gln	Ile	Met	Ala	Ile	Gln	Glu	His	Ser	Tyr	Tyr	Ala	Ser	Phe	Gly	Tyr
	385				390					395				400	
His	Val	Thr	Asn	Phe	Phe	Ala	Pro	Ser	Ser	Arg	Phe	Gly	Thr	Pro	Asp
				405				410						415	
Asp	Leu	Lys	Ser	Leu	Ile	Asp	Lys	Ala	His	Glu	Leu	Gly	Ile	Val	Val
			420					425					430		

Leu Met Asp Ile Val His Ser His Ala Ser Asn Asn Thr Leu Asp Gly  
 435 440 445  
 Leu Asn Met Phe Asp Cys Thr Asp Ser Cys Tyr Phe His Ser Gly Ala  
 450 455 460  
 Arg Gly Tyr His Trp Met Trp Asp Ser Arg Leu Phe Asn Tyr Gly Asn  
 465 470 475 480  
 Trp Glu Val Leu Arg Tyr Leu Leu Ser Asn Ala Arg Trp Trp Leu Asp  
 485 490 495  
 Ala Phe Lys Phe Asp Gly Phe Arg Phe Asp Gly Val Thr Ser Met Met  
 500 505 510  
 Tyr Ile His His Gly Leu Ser Val Gly Phe Thr Gly Asn Tyr Glu Glu  
 515 520 525  
 Tyr Phe Gly Leu Ala Thr Asp Val Asp Ala Val Val Tyr Leu Met Leu  
 530 535 540  
 Val Asn Asp Leu Ile His Gly Leu Phe Pro Asp Ala Ile Thr Ile Gly  
 545 550 555 560  
 Glu Asp Val Ser Gly Met Pro Thr Phe Cys Ile Pro Val Gln Glu Gly  
 565 570 575  
 Gly Val Gly Phe Asp Tyr Arg Leu His Met Ala Ile Ala Asp Lys Arg  
 580 585 590  
 Ile Glu Leu Leu Lys Lys Arg Asp Glu Asp Trp Arg Val Gly Asp Ile  
 595 600 605  
 Val His Thr Leu Thr Asn Arg Arg Trp Ser Glu Lys Cys Val Ser Tyr  
 610 615 620  
 Ala Glu Ser His Asp Gln Ala Leu Val Gly Asp Lys Thr Ile Ala Phe  
 625 630 635 640  
 Trp Leu Met Asp Lys Asp Met Tyr Asp Phe Met Ala Leu Asp Arg Pro  
 645 650 655  
 Ser Thr Ser Leu Ile Asp Arg Gly Ile Ala Leu His Lys Met Ile Arg  
 660 665 670  
 Leu Val Thr Met Gly Leu Gly Gly Glu Gly Tyr Leu Asn Phe Met Gly  
 675 680 685  
 Asn Glu Phe Gly His Pro Glu Trp Ile Asp Phe Pro Arg Ala Glu Gln  
 690 695 700  
 His Leu Ser Asp Gly Ser Val Ile Pro Gly Asn Gln Phe Ser Tyr Asp  
 705 710 715 720  
 Lys Cys Arg Arg Arg Phe Asp Leu Gly Asp Ala Glu Tyr Leu Arg Tyr  
 725 730 735  
 Arg Gly Leu Gln Glu Phe Asp Arg Pro Met Gln Tyr Leu Glu Asp Lys  
 740 745 750  
 Tyr Glu Phe Met Thr Ser Glu His Gln Phe Ile Ser Arg Lys Asp Glu  
 755 760 765  
 Gly Asp Arg Met Ile Val Phe Glu Lys Gly Asn Leu Val Phe Val Phe

770	775	780
Asn Phe His Trp Thr Lys Ser Tyr Ser Asp Tyr Arg Ile Ala Cys Leu 785 790 795 800		
Lys Pro Gly Lys Tyr Lys Val Ala Leu Asp Ser Asp Asp Pro Leu Phe 805 810 815		
Gly Gly Phe Gly Arg Ile Asp His Asn Ala Glu Tyr Phe Thr Phe Glu 820 825 830		
Gly Trp Tyr Asp Asp Arg Pro Arg Ser Ile Met Val Tyr Ala Pro Cys 835 840 845		
Lys Thr Ala Val Val Tyr Ala Leu Val Asp Lys Glu Glu Glu Glu 850 855 860		
Glu Glu Glu Glu Glu Glu Val Ala Ala Val Glu Glu Val Val Val Glu 865 870 875 880		
Glu Glu		

## (2) INFORMATION FOR SEQ ID NO: 16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2576 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TCATTAAAGA GGAGAAATTA ACTATGAGAG GATCTCACCA TCACCATCAC CATGGGATCT	60
TGGCTGAAAA GTCTTCTTAC AATTCCGAAT TCCGACCTTC TACAGTTGCA GCATCGGGGA	120
AAGTCCTTGT GCCTGGAACC CAGAGTGATA GCTCCTCATC CTCAACAAAC CAATTTGAGT	180
TCACTGAGAC ATCTCCAGAA AATTCCTCAG CATCAACTGA TGTAGATAGT TCAACAATGG	240
AACACGCTAG CCAGATTAAA ACTGAGAACG ATGACGTTGA GCCGTCAAGT GATCTTACAG	300
GAAGTGTTGA AGAGCTGGAT TTTGCTTCAT CACTACAAC TACAAGAAGG GTAAACTGG	360
AGGAGTCTAA AACATTAAAT ACTTCTGAAG AGACAATTAT TGATGAATCT GATAGGATCA	420
GAGAGAGGGG CATCCCTCCA CCTGGACTTG GTCAGAAGAT TTATGAAATA GACCCCTTT	480
TGACAAACTA TCGTCAACAC CTTGATTACA GGTATTCACA GTACAAGAAA CTGAGGGAGG	540
CAATTGACAA GTATGAGGGT GGTTTGGAAG CTTTTCTCG TGTTATGAA AAAATGGGTT	600
TCACTCGTAG TGCTACAGGT ATCACTTACC GTGAGTGGGC TCCTGGTGCC CAGTCAGCTG	660
CCCTCATGG AGATTTC AAC AATTGGGACG CAAATGCTGA CATTATGACT CGGAATGAAT	720
TTGGTGTCTG GGAGATTTT CTGCCAAATA ATGTGGATGG TTCTCCTGCA ATTCCTCATG	780
GGTCAGAGT GAAGATACGT ATGGACACTC CATCAGGTGT TAAGGATTCC ATTCCTGCTT	840
GGATCAACTA CTCTACAGCT TCCTGATGAA ATTCCATATA ATGGAATATA TTATGATCCA	900

CCCGAAGAGG AGAGGTATAT CTTCCAACAC CCACGGCCAA AGAAACCAAA GTCGCTGAGA	960
ATATATGAAT CTCATATTGG AATGAGTAGT CCGGAGCCTA AAATTAACCTC ATACGTGAAT	1020
TTTAGAGATG AAGTCTCTCC TCGCATAAAA AAGCTTGGGT ACAATGCGCT GCAAAATTATG	1080
GCTATTCAAG AGCATTCCTA TTATGCTAGT TTTGGTTATC ATGTCACAAA TTTTTTTGCA	1140
CCAAGCAGCC GTTTTGGAAC GCCCGACGAC CTTAAGTCTT TGATTGATAA AGCTCATGAG	1200
CTAGGAATTG TTGTTCTCAT GGACATTGTT CACAGCCATG CATCAATAA TACTTTAGAT	1260
GGACTGAACA TGTTTGACGG CACCGATAGT TGTTACTTTC ACTCTGGAGC TCGTGGTTAT	1320
CATTGGATGT GGGATTCCCG CCTTTTAAAC TATGGAACT GGGAGGTACT TAGGTATCTT	1380
CTCTCAAATG CGAGATGGTG GTTGGATGAG TTCAAATTG ATGGATTAG ATTGTATGGT	1440
GTGACATCAA TGATGTATAC TCACCACGGA TTATCGGTGG GATTCACTGG GAACTACGAG	1500
GAATACTTTG GACTCGCAAC TGATGTGGAT GCTGTTGTGT ATCTGATGCT GGTCAACGAT	1560
CTTATTCATG GGCTTTTCCC AGATGCAATT ACCATTGGTG AAGATGTTAG CGGAATGCCG	1620
ACATTTTGTA TTCCCGTTCA AGATGGGGT GTTGGCTTTG ACTATCGGCT GCATATGGCA	1680
ATTGCTGATA AATGGATTGA GTTGCTCAAG AAACGGGATG AGGATTGGAG AGTGGGTGAT	1740
ATTGTTCTATA CACTGACAAA TAGAAGATGG TCGGAAAAGT GTGTTTCATA CGCTGAAAGT	1800
CATGATCAAG CTCTAGTCGG TGATAAACT ATAGCATTCT GGCTGATGGA CAAGGATATG	1860
TATGATTTTA TGGCTCTGGA TAGACCGCCA ACATCATTA TAGATCGTGG GATAGCATTG	1920
CACAAGATGA TTAGCTTGT AACTATGGGA TTAGGAGGAG AAGGGTACCT AAATTTTCATG	1980
GGAAATGAAT TCGGCCACCC TGAGTGGATT GATTTCCCTA GGGCTGAACA ACACCTCTCT	2040
GATGACTCAG TAATTCCCGG AAACCAATTC AGTTATGATA AATGCAGACG GAGATTTGAC	2100
CTGGGAGATG CAGAATATTT AAGATACCGT GGGTTGCAAG AATTTGACCG GGCTATGCAG	2160
TATCTTGAAG ATAAATATGA GTTTATGACT TCAGAACACC AGTTCATATC ACGAAAGGAT	2220
GAAGGAGATA GGATGATTGT ATTTGAAAAA GGAAACCTAG TTTTGTCTT TAATTTTCAC	2280
TGGCAAAAAA GCTATTTCAGA CTATCGCATA GGCTGCCTGA AGCCTGGAAA ATACAAGGTT	2340
GCCTTGACT CAGATGATCC ACTTTTGGT GGCTTCGGGA GAATTGATCA TAATGCCGAA	2400
TATTTACCT TTGAAGGATG GTATGATGAT CGTCCTCGTT CAATTATGGT GTATGCACCT	2460
TGTAGAACAG CAGTGGTCTA TGCACTAGTA GACAAAGAAG AAGAAGAAGA AGAAGAAGAA	2520
GAAGAAGTAG CAGTAGTAGA AGAAGTAGTA GTAGAAGAAG AATGAACGAA CTTGTG	2576

## (2) INFORMATION FOR SEQ ID NO: 17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2529 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GGATGCTAAT GTTCTGTAT TCTTGA AAAA GCACTCTCTT TCACGGAAGA TCTTGGCTGA	60
AAAGTCTTCT TACAATTCGG AATCCCAGCC TTCTACAGTT GCAGCATCGG GGAAGTCCT	120
TGTGCCTGGA AYCCAGAGTG ATAGCTCCTC ATCCTCAACA GACCAATTG AGTTCAGTGA	180
GACATCTCCA GAAAATTTCC CAGCATCAAC TGATGTAGAT AGTTCAACAA TGGAACACGC	240
TAGCCAGATT AAAACTGAGA ACGATGACGT TGAGCCGTCA AGTGATCTTA CAGGAAGTGT	300
TGAAGAGCTG GATTTTGCTT CATCACTACA ACTACAAGAA GGTGTTAAAC TGGAGGAGTC	360
TAAACATTA AATACTTCTG AAGAGACAAT TATTGATGAA TCTGATAGGA TCAGAGAGAG	420
GGGCATCCCT CCACCTGGAC TTGGTCAGAA GATTTATGAA ATAGACCCCC TTTTGACAAA	480
CTATCGTCAA CACCTTGATT ACAGGTATTC ACAGTACAAG AAAGTGGGG AGGCAATTGA	540
CAAGTATGAG GGTGGTTTGG AAGCTTTTTT TCGTGGTTAT GAAAAAATGG GTTTCAGTCG	600
TAGTGCTACA GGTATCACTT ACCGTGAGTG GGCTCCTGGT GCCCAGTCAG CTGCCCTCAT	660
TGGAGATTTC AACAATTGGG ACGCAAATGC TGACATTATG ACTCGGAATG AATTGGTGT	720
CTGGGAGATT TTTCTGCCAA ATAATGTGGA TGGTTCTCCT GCAATTCCTC ATGGGTCCAG	780
AGTGAAGATA CGYATGGACA CTCCATCAGG TGTTAAGGAT TCCATTCTCG CTTGGATCAA	840
CTACTCTTTA CAGCTTCCTG ATGAAATTC ATATAATGGA ATATATTATG ATCCACCCGA	900
AGAGGAGAGG TATRTCTTCC AACACCCACG GCCAAAGAAA CCAAAGTCGC TGAGAATATA	960
TGAATCTCAT ATTGGAATGA GTAGTCCGGA GCCTAAAATT AACTCATAAG TGAATTTTAG	1020
AGATGAAGTT CTTCTCGCA TAAAAAASCT TGGGTACAAT GCGGTGCAAA TTATGGCTAT	1080
TCAAGAGCAT TCTTATTATG CTAGTTTTGG TTATCATGTC ACAAATTTTT TTGCACCAAG	1140
CAGCCGTTTT GGAACGCCCG ACGACCTTAA GTCTTTGATT GATAAAGCTC ATGAGCTAGG	1200
AATTGTGTGT CTATGGACA TTGTTCACAG CCATGCATCA AATAATACTT TAGATGGACT	1260
GAACATGTTT GACGGCACAG ATAGTTGTTA CTTTCACTCT GGAGCTCGTG GTTATCATTG	1320
GATGTGGGAT TCCCGCCTCT TTAATATGGA AAAGTGGGAG GTACTTAGGT ATCTTCTCTC	1380
AAATGCGAGA TGGTGGTTGG ATGAGTTCAA ATTTGATGGA TTAGATTG ATGGTGTGAC	1440
ATCAATGATG TATACTCACC ACGGATTATC GGTGGGATTC ACTGGGAAGT ACGAGGAATA	1500
CTTGACTC GCAACTGATG TGGATGCTGT TGTGTATCTG ATGCTGGTCA ACGATCTTAT	1560
TCACGGGCTT TTCCAGATG CAATTACCAT TGGTGAAGAT GTTAGCGGAA TGCCGACATT	1620
TTGTATTCCC GTTCAAGATG GGGGTGTTGG CTTTGACTAT CGGCTGCATA TGGCAATTGC	1680
TGATAAATGG ATTGAGTTGC TCAAGAAACG GGATGAGGAT TGGAGAGTGG GTGATATTGT	1740
TCATACACTG ACAAATAGAA GATGGTCGGA AAAGTGTGTT TCATMCGCTG AAAGTCATGA	1800



TCAAGCTCTA GTCGGTGATA AAACCTATAGC ATYCTGGCTG ATGGACAAGG ATATGTATGA	1860
TTTTATGGCT CTGGATAGAC CGYCAACAYC ATTAATAGAT CGTGGGATAG CATTGCACAA	1920
GATGATTAGG CTTGTAAC TA TGGGATTAGG AGGAGAAGGG TACCTAAATT TCATGGGAAA	1980
TGAATTCGGC CACCCTGAGT GGATTGATTT CCCTAGGGCT GARCAACACC TCTCTGATGG	2040
CTCAGTAATT CCCGAAACC AATTTCAGTTA TGATAAATGC AGACGGAGAT TTGACCTGGG	2100
AGATGCAGAA TATTTAAGAT ACCATGGGTT GCAAGAATTT GACCGGGCTA TGCAGTATCT	2160
TGAAGATAAA TATGAGTTTA TGACTTCAGA ACACCAGTTC ATATCACGAA AGGATGAAGG	2220
AGATAGGATG ATTGTATTG AAARAGGAAA CCTAGTTTTT GTCTTTAATT TTCACTGGAC	2280
AAATAGCTAT TCAGACTATC GCATAGGCTG CCTGAAGCCT GGAAAATACA AGGTTGGCTT	2340
GGACTCAGAT GATCCACTTT TTGGTGGCTT CGGGAGAATT GATCATAATG CCGAATATTT	2400
CACCTCTGAA GGATCGTATG ATGATCGTCC TCGTTCAATT ATGGTGTATG CACCTAGTAG	2460
AACAGCAGTG GTCTATGCAC TAGTAGACAA ANTAGAAGNA GAAGAAGAAG AAGAANCCGN	2520
NGAAGAATT	2529

## (2) INFORMATION FOR SEQ ID NO: 18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3231 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GATTTAATAC GACTCACTAT AGGGATTTTT TTTTTTTTTT TTTTAAAAAC CTCCTCCACT	60
CAGTCTTGGG ATCTCTCTCT CTCTTCACGC TTCTCTTGGG GCCTTGAAC T CAGCAATTTG	120
ACACTCAGTT AGTTCACTC CTATCACTCA TCAGATCTCT ATTTTTTCTC TTAATTCCAA	180
CCAAGGAATG AATTAAAAGA TTAGATTTGA AGGAGAGAAG AAGAAAGATG GTGTATACAC	240
TCTCTGGAGT TCGTTTTCTT ACTGTTCCAT CAGTGACAA ATCTAATGGA TTCAGCAGTA	300
ATGGTGATCG GAGGAATGCT AATGTTTCTG TATTCTTGAA AAAGCACTCT CTTTCACGGA	360
AGATCTTGGC TGAAAAGTCT TCTTACGATT CCGAATCCCG ACCTTCTACA GTTGCAGCAT	420
CGGGGAAAGT CTTGTACCT GGAATCCAGA GTGATAGCTC CTCATCCTCA ACAGACCAAT	480
TTGAGTTCAC TGAGACAGCT CCAGAAAATT CCCAGCATC AACTGATGTG GATAGTTCAA	540
CAATGGAACA CGCTAGCCAG ATTAAAACTG AGAACGATGA CGTTGAGCCG TCAAGTGATC	600
TTACAGGAAG TGTTGAAGAG TTGGATTTTG CTTCATCACT ACAACTACAA GAAGGTGGTA	660
AACTGGAGGA GTCTAAAACA TTAAATACTT CTGAAGAGAC AATTATTGAT GAATCTGATA	720
GGATCAGAGA GAGGGGCATC CCTCCACCTG GACTTGGTCA GAAGATTTAT GAAATAGACC	780

CCCTTTTGAC	AACTATCGT	CAACACCTTG	ATTACAGGTA	TTCACAGTAC	AAGAAAATGA	840
GGGAGGCAAT	TGACAAGTAT	GAGGGTGGTT	TGGAAGCTTT	TTCTCGTGGT	TATGAAAAAA	900
TGGGTTTCAC	TCGTAGTGCT	ACAGGTATCA	CTTACCGTGA	GTGGGCTCCT	GGTGGCCAGT	960
CAGCTGCTCT	CATTGGAGAT	TTCAACAATT	GGGACGCAA	TGCTGACATT	ATGACTCGGA	1020
ATGAATTG	TGTCTGGGAG	ATTTTCTG	CAAATAATGT	GGATGGTTCT	CCTGCAATTC	1080
CTCATGGGTC	CAGAGTGAAG	ATACGCATGG	ACACTTCATC	AGGTGTTAAG	GATTCCATTC	1140
CTGCTTGGAT	CAACTACTCT	TTACAGCTTC	CTGATGAAAT	TCCATATAAT	GGAATATATT	1200
ATGATCCACC	CGAAGAGGAG	AGGTATGTCT	TCCAACACCC	ACGGCCAAAG	AAACCAAAGT	1260
CGCTGAGAAT	ATATGAATCT	CATATTGGAA	TGAGTAGTCC	GGAGCCTAAA	ATTAACATCAT	1320
ACGTGAATTT	TAGAGATGAA	GTTCTTCCTC	GCATAAAAAA	CCTTGGGTAC	AATGCGGTGC	1380
AAATTATGGC	TATTCAAGAG	CATTCTTATT	ATGCTAGTTT	TGGTTATCAT	GTCACAAATT	1440
TTTTTGCACC	AAGCAGCCGT	TTTGAACGC	CCGACGACCT	TAAGCTTTTG	ATTGATAAAG	1500
CTCATGAGCT	AGGAATTGTT	GTTCTCATGG	ACATTGTTCA	CAGCCATGCA	TCAAATAATA	1560
CTTTAGATGG	ACTGAACATG	TTTGACGGCA	CAGATAGTTG	TTACTTTCAC	TCTGGAGCTC	1620
GTGGTTATCA	TTGGATGTGG	GATTCCCGCC	TCTTTAACTA	TGGAAACTGG	GAGGTACTTA	1680
GGTATCTTCT	CTCAAATGCG	AGATGGTGGT	TGGATGAGTG	CAAATTTGRT	GGATTTAGAT	1740
TTGATGGTGT	GACATCAATG	ATGTATACTC	ACCACGGATT	ATCGGTGGGA	TTCCTGGGA	1800
ACTACGAGGA	ATACTTTGGA	CTCGCAACTG	ATGTRGATGC	TGCCGTGTAT	CTGATGCTGG	1860
CCAACGATCT	TATTATGGG	CTTTTCCAG	ATGCAATTAC	CATTGGTGAA	GATGTTAGCG	1920
GAATGCCGAC	ATTTTGTATT	CCCGTTCAAG	ATGGGGTGT	TGGCTTTGAC	TATCGGCTGC	1980
ATATGGCAAT	TGCTGATAAA	TGGATTGAGT	TGCTCAAGAA	ACGGGATGAG	GATTGGAGAG	2040
TGGGTGATAT	TGTTCATACA	CTGACAAATA	GAAGATGGTC	GGAAAAGTGT	GTTTCATACG	2100
CTGAAAGTCA	TGATCAAGCT	CTAGTCGGTG	ATAAACTAT	AGCATTCTGG	CTGATGGACA	2160
AGGATATGTA	TGATTTTATG	GCTTTGGATA	GACCGTCAAC	ATCATTAAATA	GATCGTGGGA	2220
TAGCATTGCA	CAAGATGATT	AGGCTTGTA	CTATGGGATT	AGGAGGAGAA	GGGTACCTAA	2280
ATTTATGGG	AAATGAATTC	GGCCACCTG	AGTGGATTGA	TTCCCTAGG	GCTGAACAAC	2340
ACCTCTCTGA	TGGCTCAGTA	ATCCCGGAA	ACCAATTGAG	TTATGATAAA	TGCAGACGGA	2400
GATTTGACCT	GGGAGATGCA	GAATATTTAA	GATACCGTGG	GTTGCAAGAA	TTTGACCGGG	2460
CTATGCAGTA	TCTGAAGAT	AAATATGAGT	TTATGACTTC	AGAACACCAG	TTCATATCAC	2520
GAAAGGATGA	AGGAGATAGG	ATGATTGTAT	TTGAAAAAGG	AAACCTAGTT	TTTGCTTTTA	2580
ATTTTCACTG	GACAAAAAGC	TATTCAGACT	ATCGCATAGG	CTGGCTGAAG	CCTGGAATAAT	2640
ACAAGGTGTC	CTTGACTCA	GATGATCCAC	TTTTTGGTGG	CTTCGGGAGA	ATTGATCATA	2700

ATGCCGAATG	TTTCACCTTT	GAAGGATGGT	ATGATGATCG	TCCTCGTTCA	ATTATGGTGT	2760
ATGCACCTAG	TAGAACAGCA	GTGGTCTATG	CACTAGTAGA	CAAAGAAGAA	GAAGAAGAAG	2820
AAGTAGCAGT	AGTAGAAGAA	GTAGTAGTAG	AAGAAGAATG	AACGAACCTG	TGATCGCGTT	2880
GAAAGATTTG	AACGCTACAT	AGAGCTTCTT	GACGTATCTG	GCAATATTGC	ATCAGTCTTG	2940
GCGGAATTTT	ATGTGACAAA	AGGTTTGCAA	TTCTTTCCAC	TATTAGTAGT	GCAACGATAT	3000
ACGCAGAGAT	GAAGTGCTGA	ACAAACATAT	GTAAAAATCGA	TGAATTTATG	TCGAATGCTG	3060
GGACGGGCTT	CAGCAGGTTT	TGCTTAGTGA	GTTCTGTAAA	TTGTCATCTC	TTTANATGTA	3120
CAGCCCCTA	GAAATCAATT	ATGTGAGACC	TAAAAAACAA	TAACCATAAA	ATGGAAATAG	3180
TGCTGATCTA	ATGATGTTTT	AANCCNNNNA	AAAAAAAAAA	AAAAACTCGA	G	3231

## (2) INFORMATION FOR SEQ ID NO: 19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2578 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TCATTAAAGA	GGAGAAATTA	ACTATGAGAG	GATCTCACCA	TCACCATCAC	CATGGGATCT	60
TGGCTGAAAA	GTCTTCTTAC	AATTCGGAAT	TCCGACCTTC	TACAGTTGCA	GCATCGGGGA	120
AAGTCCTTGT	GCCTGGAACC	CAGAGTGATA	GCTCCTCATC	CTCAACAAAC	CAATTTGAGT	180
TCACTGAGAC	ATCTCCAGAA	AATTCCCCAG	CATCAACTGA	TGTAGATAGT	TCAACAATGG	240
AACACGCTAG	CCAGATTAAG	ACTGAGAACG	ATGACGTTGA	GCCGTCAAGT	GATCTTACAG	300
GAAGTGTTGA	AGAGCTGGAT	TTTGCTTCAT	CACTACAAC	ACAAGAAGGT	GGTAACTGG	360
AGGAGTCTAA	AACATTAAAT	ACTTCTGAAG	AGACAATTAT	TGATGAATCT	GATAGGATCA	420
GAGAGAGGGG	CATCCCTCCA	CCTGGACTTG	GTCAGAAGAT	TTATGAAATA	GACCCCTTTT	480
TGACAAACTA	TCGTCAACAC	CTTGATTACA	GGTATTCACA	GTACAAGAAA	CTGAGGGAGG	540
CAATTGACAA	GTATGAGGGT	GGTTTGGAAG	CTTTTCTCG	TGTTTATGAA	AAAATGGGTT	600
TCACTCGTAG	TGCTACAGGT	ATCACTTACC	GTGAGTGGGC	TCCTGGTGCC	CAGTCAGCTG	660
CCCTCATTGG	AGATTTCAAC	AATTGGGACG	CAAATGCTGA	CATTATGACT	CGGAATGAAT	720
TTGGTGTCTG	GGAGATTTTT	CTGCCAAATA	ATGTGGATGG	TTCTCCTGCA	ATTCCTCATG	780
GGTCCAGAGT	GAAGATACGT	ATGGACACTC	CATCAGGTGT	TAAGGATTCC	ATTCCTGCTT	840
GGATCAACTA	CTCTTCACAG	CTTCCTGATG	AAATCCATA	TAATGGAATA	TATTATGATC	900
CACCCGAAGA	GGAGAGGTAT	ATCTTCCAAC	ACCCACGGCC	AAAGAAACCA	AAGTCGCTGA	960
GAATATATGA	ATCTCATATT	GGAATGAGTA	GTCCGGAGCC	TAAATTAAC	TCATACGTGA	1020

ATTTTAGAGA	TGAAGTTCTT	CCTCGCATAA	AAAAGCTTGG	GTACAATGCG	GTGCAAATTA	1080
TGGCTATTCA	AGAGCATTCT	TATTATGCTA	GTTTTGGTTA	TCATGTCACA	AATTTTTTTG	1140
CACCAAGCAG	CCGTTTTGGA	ACGCCCCGACG	ACCTTAAGTC	TTTGATTGAT	AAAGCTCATG	1200
AGCTAGGAAT	TGTTGTTCTC	ATGGACATTG	TTCACAGCCA	TGCATCAAAT	AATACTTTAG	1260
ATGGACTGAA	CATGTTTGAC	GGCACCAGATA	GTTGTTACTT	TCACTCTGGA	GCTCGTGGTT	1320
ATCATTGGAT	GTGGGATTCC	CGCCTTTTTA	ACTATGGAAA	CTGGGAGGTA	CTTAGGTATC	1380
TTCTCTCAAA	TGCAGATGG	TGGTTGGATG	AGTTCAAATT	TGATGGATTT	AGATTTGATG	1440
GTGTGACATC	AATGATGTAT	ACTCACCACG	GATTATCGGT	GGGATTCACT	GGGAACACG	1500
AGGAATACTT	TGGACTCGCA	ACTGATGTGG	ATGCTGTTGT	GTATCTGATG	CTGGTCAACG	1560
ATCTTATTCA	TGGGCTTTTC	CCAGATGCAA	TTACCATTGG	TGAAGATGTT	AGCGGAATGC	1620
CGACATTTTG	TATTCCTGTT	CAAGATGGGG	GTGTTGGCTT	TGACTATCGG	CTGCATATGG	1680
CAATTGCTGA	TAAATGGATT	GAGTGTCTCA	AGAAACGGGA	TGAGGATTGG	AGAGTGGGTG	1740
ATATTGTTCA	TACACTGACA	AATAGAAGAT	GGTCGGAAAA	GTGTGTTTCA	TACGCTGAAA	1800
GTCATGATCA	AGCTCTAGTC	GGTGATAAAA	CTATAGCATT	CTGGCTGATG	GACAAGGATA	1860
TGTATGATTT	TATGGCTCTG	GATAGACCGC	CAACATCATT	AATAGATCGT	GGGATAGCAT	1920
TGCACAAGAT	GATTAGGCTT	GTAACATATG	GATTAGGAGG	AGAAGGGTAC	CTAAATTTCA	1980
TGGGAAATGA	ATTGCGCCAC	CCTGAGTGGA	TTGATTTCCT	TAGGGCTGAA	CAACACCTCT	2040
CTGATGACTC	AGTAATTCCT	GGAAACCAAT	TCAGTTATGA	TAAATGCAGA	CGGAGATTTG	2100
ACCTGGGAGA	TGCAGAATAT	TTAAGATACC	GTGGGTTGCA	AGAATTTGAC	CGGGCTATGC	2160
AGTATCTTGA	AGATAAATAT	GAGTTTATGA	CTTCAGAACA	CCAGTTTATA	TCACGAAAGG	2220
ATGAAGGAGA	TAGGATGATT	GTATTTGAAA	AAGGAAACCT	AGTTTTTGTC	TTTAATTTTC	2280
ACTGGACAAA	AAGCTATTCA	GAATATCGCA	TAGGCTGCCT	GAAGCCTGGA	AAATACAAGG	2340
TTGCCTTGGA	CTCAGATGAT	CCACTTTTTG	GTGGCTTCGG	GAGAATTGAT	CATAATGCCG	2400
AATATTTTAC	CTTTGAAGGA	TGGTATGATG	ATCGTCCTCG	TTCAATTATG	GTGTATGCAC	2460
CTTGTAGAAC	AGCAGTGGTC	TATGCACTAG	TAGACAAAGA	AGAAGAAGAA	GAAGAAGAAG	2520
AAGAAGAAGT	AGCAGTAGTA	GAAGAAGTAG	TAGTAGAAGA	AGAATGAACG	AACCTTGTG	2578

## (2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AATTTYATGG GNAAYGARTT YGG